

Machine Learning Final Project



Prediksi Nilai Rapor Mahasiswa





Group [7] Role's...



[Alvinhugo Zuhdi]

EDA



[Syahrul Azka]

Modelling



[Syifa Ghaisani]

PPT



AI-ML Weekly Class

```
function filterStudies({ studies, filterByOrg = false, filterByTopic = false, filterByYear = false, filterByOrganization = false }) {  
  return studies.filter(study => {  
    if (filterByOrg) {  
      return study.organization === filterByOrg;  
    }  
    if (filterByTopic) {  
      return study.topic === filterByTopic;  
    }  
    if (filterByYear) {  
      return study.year === filterByYear;  
    }  
    if (filterByOrganization) {  
      return study.organization === filterByOrganization;  
    }  
    return true;  
  });  
}
```

Case



1



Case

Seorang rektor akan menerapkan kebijakan pembelajaran yang efektif. Namun dia perlu adanya bantuan insight dari seorang machine learning engineer, agar memiliki kebijakan yang bagus. Anda akan diberikan beberapa informasi tentang mahasiswa seperti: jumlah mata kuliah, rata-rata belajar per hari, dan nilai yang diperoleh mahasiswa. Dengan menggunakan informasi ini, Sebagai seorang machine learning engineer, apa yang akan kalian lakukan supaya dapat memberikan insight kepada rektor?

2

3

4

Strategy



1

1. Describe



Dataset yang diberikan sudah sangat bersih, tidak ada outlier, missing value, duplicate. Data yang kami miliki berbentuk numerik

2. EDA



- Univariate Analysis
- Bivariate Analysis
- Multivariate Analysis

3. Feature Engineering



Data yang kami miliki tidak memiliki outlier, missing value, duplicate

4. Model + Evaluation



- Data dibagi menjadi train & test (70;30)
- Inisialisasi model regresi linear
- Latih model menggunakan data train
- Prediksi nilai Y
- Evaluasi model dengan membandingkan nilai Y predict dengan Y actual

2

3

4

Describe Data



1

Dataset

- Volume Data:
100 rows and 3 columns

```
df.shape
```

```
(100, 3)
```

Type Data

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 100 entries, 0 to 99  
Data columns (total 3 columns):  
#   Column          Non-Null Count  Dtype  
---  ---  
0   number_courses  100 non-null   int64  
1   time_study      100 non-null   float64  
2   Marks           100 non-null   float64  
dtypes: float64(2), int64(1)  
memory usage: 2.5 KB
```

2

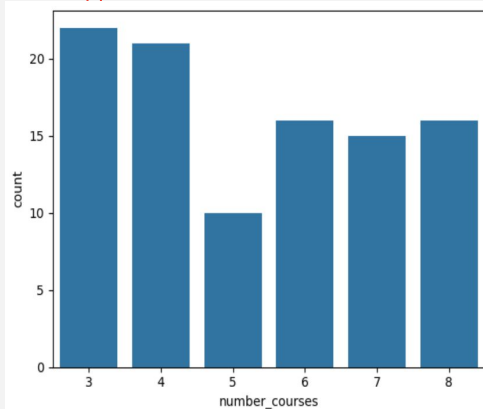
Feature

- Independen : number_courses, time_study
- Dependen/Target : Marks

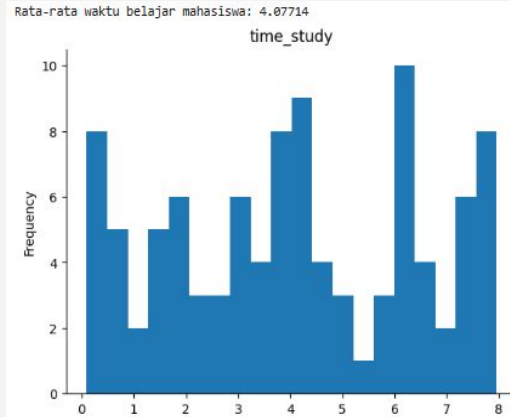
3

4

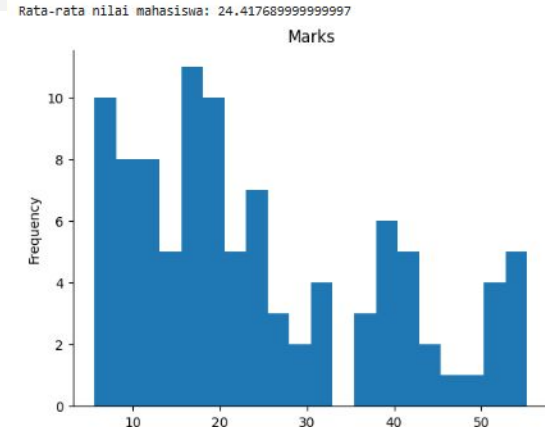
EDA Univariate Analysis



Terlihat bahwa jumlah mahasiswa yang mengambil 3 mata kuliah adalah yang tertinggi dari semua kategori jumlah mata kuliah.



Terlihat bahwa frekuensi rata-rata waktu belajar mahasiswa terbesar berada di sekitar 6 jam.



Terlihat bahwa frekuensi rata-rata nilai mahasiswa terbesar berada di sekitar 15.

1

2

3

4

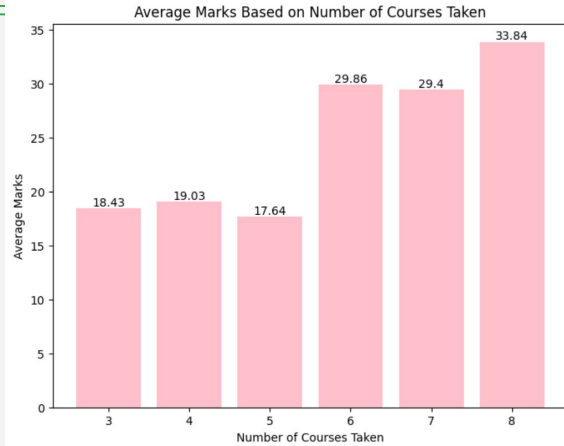


EDA

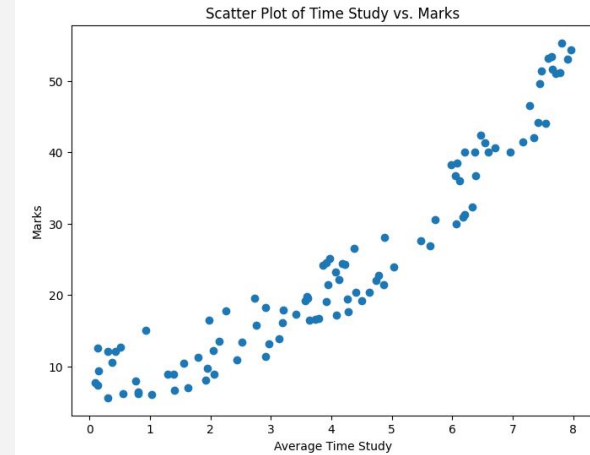
Bivariate Analysis



1



Terlihat bahwa rata-rata nilai tertinggi terjadi pada mahasiswa yang mengambil 8 mata kuliah



Rata-rata lama waktu belajar mahasiswa punya pengaruh yang signifikan terhadap nilai mahasiswa. Semakin banyaknya waktu belajar, maka semakin tinggi juga nilai yang mahasiswa dapatkan.

2

3

4



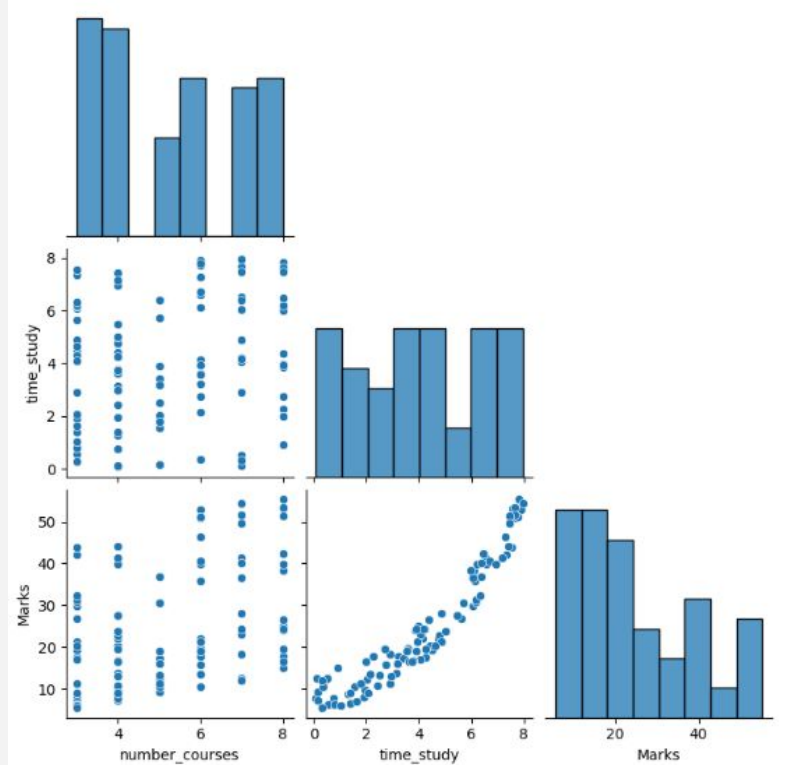
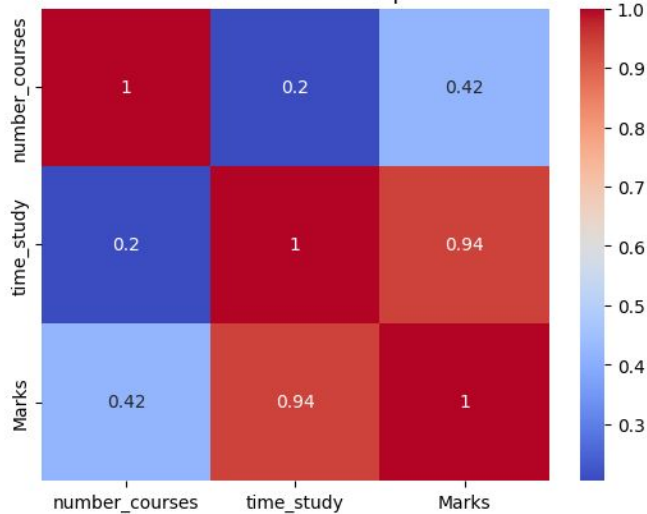
EDA

Multivariate Analysis



1

Correlation Heatmap



2

3

4

Feature Engineering



1

Filled with Mean



Missing Value

Duplicated



Remove it

Mengubah Type Data



Inconsistent Data

Label Kategorikal



One-hot encoding

Data Discretization



Feature Numerical

Outlier



Standardization (Feature Normalization)

2

3

4

Model Building



1



10%

[Point] Tantangan

80%

[Point] Strategy

80%

[Point] Solusi

2

3

4

Model Evaluation



1

MAE

3,03

MSE

13,10

R Square

0,94

```
mse = mean_squared_error(y_test, y_pred)
mae = mean_absolute_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
```

```
print("Mean Squared Error (MSE):", mse)
print("Mean Absolute Error (MAE):", mae)
print("Nilai R Square:", r2)
```

```
Mean Squared Error (MSE): 13.104475429070757
Mean Absolute Error (MAE): 3.0284743686391673
Nilai R Square: 0.9417542675032117
```

2

3

4



Summary



1



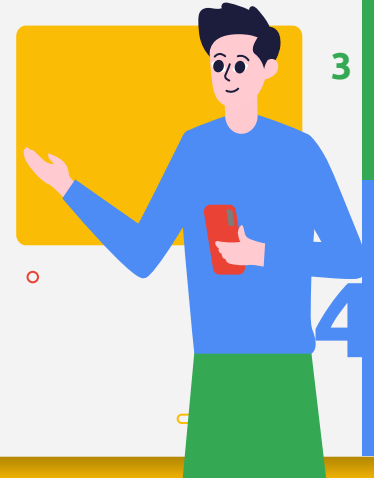
- Waktu belajar mahasiswa memiliki pengaruh yang sangat besar terhadap nilai yang diperoleh. Nilai korelasi variabel 'time_study' terhadap variabel 'marks' sebesar 0.94 menunjukkan hubungan positif yang kuat antara waktu belajar dan nilai.
- Banyaknya mata kuliah yang diambil memiliki pengaruh terhadap nilai yang diperoleh. Nilai korelasi sebesar 0.42 menunjukkan hubungan positif antara jumlah mata kuliah dan nilai. Meskipun korelasinya tidak sangat kuat, tetapi tetap menunjukkan bahwa semakin banyak mata kuliah yang diambil, kemungkinan nilai yang lebih tinggi dapat diperoleh.
- Nilai MSE (13.10) dan MAE (3.02) yang cukup rendah.
- Nilai R Square yang tinggi (0.93) menunjukkan bahwa model kita dapat menjelaskan sebagian besar variasi dalam nilai rata-rata mahasiswa, atau dengan kata lain sekitar 93% variasi dalam nilai dapat dijelaskan oleh fitur-fitur yang digunakan dalam model.



2



3



4



Insight & Recommendation



1

- Karena waktu belajar mahasiswa memiliki pengaruh yang sangat besar terhadap nilai yang diperoleh, penting bagi universitas untuk memastikan bahwa mahasiswa memiliki akses dan kesempatan untuk belajar dengan waktu yang memadai. Rektor dapat mempertimbangkan untuk mengimplementasikan kebijakan atau program yang mendukung pembelajaran aktif di luar kelas, seperti kelas tambahan, bimbingan akademik, atau sumber daya belajar online selain mata kuliah wajib di dalam kelas.
- Meskipun korelasi antara jumlah mata kuliah yang diambil dan nilai tidak sangat kuat, masih ada hubungan positif antara keduanya. Oleh karena itu, rektor dapat mempertimbangkan untuk memastikan bahwa jumlah mata kuliah yang diambil oleh mahasiswa sesuai dengan kapasitas dan kemampuan mereka. Ini dapat dilakukan dengan memastikan adanya bimbingan akademik yang memadai dan memperhatikan beban kuliah mahasiswa agar tidak terlalu berat.
- Model masih memiliki keterbatasan dalam memprediksi nilai mahasiswa. Karena fitur yang digunakan hanya banyaknya jumlah mata kuliah yang diambil dan waktu belajar mahasiswa. Rektor dapat melakukan analisis lebih lanjut terhadap faktor-faktor lainnya yang menyebabkan nilai mahasiswa menjadi rendah atau tinggi.

2

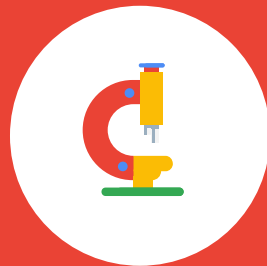
3

4





ML Team Engineer Group [7]



Thank You